



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

# A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway is Associated with Major Depressive Disorder

### Citation for published version:

Zeng, Y, Navarro, P, Fernandez-pujals, AM, Hall, LS, Clarke, T, Thomson, PA, Smith, BH, Hocking, LJ, Padmanabhan, S, Hayward, C, Macintyre, DJ, Wray, NR, Deary, I, Porteous, DJ, Haley, CS & McIntosh, AM 2017, 'A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway is Associated with Major Depressive Disorder' *Biological Psychiatry*, vol. 81, no. 4, pp. 336-346. DOI: 10.1016/j.biopsych.2016.04.017

### Digital Object Identifier (DOI):

[10.1016/j.biopsych.2016.04.017](https://doi.org/10.1016/j.biopsych.2016.04.017)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

Biological Psychiatry

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

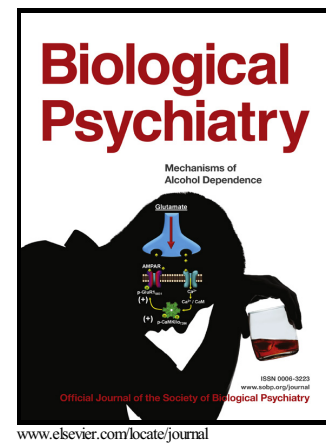
### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway is Associated with Major Depressive Disorder pathway and regional heritability analyses in MDD

Yanni Zeng, Pau Navarro, Ana M. Fernandez-Pujals, Lynsey S. Hall, Toni-Kim Clarke, Pippa A. Thomson, Blair H. Smith, Lynne J. Hocking, Sandosh Padmanabhan, Caroline Hayward, Donald J. MacIntyre, Naomi R. Wray, Ian J Deary, David J Porteous, Chris S. Haley, Andrew M. McIntosh



PII: S0006-3223(16)32332-0  
DOI: <http://dx.doi.org/10.1016/j.biopsych.2016.04.017>  
Reference: BPS12872

To appear in: *Biological Psychiatry*

Received date: 7 December 2015  
Revised date: 20 April 2016  
Accepted date: 21 April 2016

Cite this article as: Yanni Zeng, Pau Navarro, Ana M. Fernandez-Pujals, Lynsey S. Hall, Toni-Kim Clarke, Pippa A. Thomson, Blair H. Smith, Lynne J. Hocking, Sandosh Padmanabhan, Caroline Hayward, Donald J. MacIntyre, Naomi R. Wray, Ian J Deary, David J Porteous, Chris S. Haley and Andrew M. McIntosh, A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway is Associated with Major Depressive Disorder pathway and regional heritability analyses in MDD, *Biological Psychiatry*, <http://dx.doi.org/10.1016/j.biopsych.2016.04.017>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway is Associated with Major Depressive Disorder

Short title: pathway and regional heritability analyses in MDD

Yanni Zeng<sup>1\*</sup>, Pau Navarro<sup>2</sup>, Ana M. Fernandez-Pujals<sup>1</sup>, Lynsey S. Hall<sup>1</sup>, Toni-Kim Clarke<sup>1</sup>, Pippa A. Thomson<sup>3,4</sup>, Blair H. Smith<sup>5,6</sup>, Lynne J. Hocking<sup>6,7</sup>, Sandosh Padmanabhan<sup>6,8</sup>, Caroline Hayward<sup>6,9</sup>, Donald J. MacIntyre<sup>1</sup>, Naomi R. Wray<sup>10</sup>, Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium<sup>^</sup>, Ian J Deary<sup>3,11,12</sup>, David J Porteous<sup>3,4,12</sup>, Chris S. Haley<sup>2,13</sup> and Andrew M. McIntosh<sup>1,3,12</sup>

## Affiliations:

<sup>1</sup> Division of Psychiatry, University of Edinburgh, Edinburgh, UK

<sup>2</sup> MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

<sup>3</sup> Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom

<sup>4</sup> Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh

<sup>5</sup> Division of Population Health Sciences, University of Dundee, Dundee, UK

<sup>6</sup> Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

<sup>7</sup> Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

<sup>8</sup> Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

<sup>9</sup> MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

<sup>10</sup> Queensland Brain Institute, University of Queensland, St Lucia, Queensland, Australia

<sup>11</sup> Department of Psychology, University of Edinburgh, UK

<sup>12</sup> Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

<sup>13</sup> The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, UK

\* Corresponding author

Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK EH10 5HF; T: +44 (131) 537 6687

E: Y.Zeng-6@sms.ed.ac.uk (YNZ)

**Key words:** MDD; pathway analysis; regional heritability; NETRIN1; DCC; Polygenic risk score

**Abstract**

Background: genome-wide association studies of Major Depressive Disorder (MDD) have identified few significant associations. Testing the aggregation of genetic variants in particular biological pathways may be more powerful. Regional heritability analysis can be used to detect genomic regions contributing to disease risk.

Methods: we integrated pathway analysis and multi-level regional heritability analyses in a pipeline designed to identify MDD-associated pathways. The pipeline was applied to two independent GWAS studies (GS:SFHS, N=6,455) and PGC:MDD (N=18,755). A polygenic risk score (PRS) composed of SNPs from the pathway most consistently associated with MDD was created and its accuracy to predict MDD using AUC, logistic regression and linear mixed model (LMM) analysis was tested.

Results: In GS:SFHS, four pathways were significantly associated with MDD, and two of these explained a significant amount of pathway-level regional heritability. In PGC:MDD, one pathway was significantly associated with MDD. Pathway-level regional heritability was significant in this pathway in one subset of PGC:MDD. For both samples the regional heritabilities were further localized to the gene and sub-region levels. The NETRIN1 signaling pathway showed the most consistent association with MDD across the two samples. PRSs from this pathway showed

competitive predictive accuracy when compared with the whole-genome PRSs when using AUC statistics, logistic regression and LMM.

Conclusions: these post-GWAS analyses highlight the value of combining multiple methods on multiple GWAS data for the identification of risk pathways for MDD. The NETRIN1 signaling pathway is identified as a candidate pathway for MDD, and should be explored in further large population studies.

## Introduction

Major Depressive Disorder (MDD) contributes 8.2% of the global burden of disease(1). Twin studies have estimated the narrow sense heritability of MDD to be 37%, confirming the involvement of genetic factors in MDD(2). However, published GWAS of MDD have only detected two loci associated with recurrent MDD at genome-wide significance in a study of Chinese women(3,4), despite of the success of GWAS for other psychiatric disorders(5). Additional methods for detecting the aggregate effects of sub-genome-wide significant risk variants are required to better extract information from available data.

Two lessons relevant to MDD can be learned from previous studies of polygenic diseases. First, disease-associated variants are enriched in functionally annotated regions of the genome(6). Second, the small signals from individual genetic variants

contrast with the stronger signals from individual pathways(7). Using GWAS summary statistics, a recent PGC study identified disease-specific and shared pathways across multiple psychiatric diseases(8). These findings suggest that the cumulative effects from single variants converge on biological pathways and the pathways themselves may be more tractable targets for GWAS. In the current study, we sought to test whether the aggregate effects of low-penetrance variants become detectable at the pathway level in MDD.

Various approaches have been developed to identify the association between pathways and phenotype. Methods designed to be applied to raw genotypes or summary statistics of GWAS have been developed(9). The optimal method should depend on the data type available and the research goals. For instance, a pathway-based study demonstrated the feasibility of identifying pathways and genes associated with schizophrenia using analytical methods that are designed for different data types (raw genotypes and GWAS summary statistics) in three independent samples (10).

Genomic Restricted Maximum Likelihood (GREML) analysis methods (11) can be used to estimate the additive variance contributed from all the genotyped SNPs using linear mixed modeling (LMM). They may also be adapted to further partition the variance components by functionally annotated SNP categories (11-13). Generally, for

polygenic traits the proportion of the phenotypic variance explained by the SNPs was proportional to the number of SNPs involved(14). However, genomic regions that explain more heritability than expected when accounting for the number of SNPs they contain have been uncovered. These regions usually overlap with regulatory, genic or conserved regions of the human genome(12,13,15,16). The locally enriched heritability is defined as regional heritability. Regional heritability analysis can be applied to identify genomic regions that contribute a significant proportion of heritability as an alternative association test(17). Using this method, a recent study identified more genomic regions reaching the suggestive level of significance than GWAS, suggesting that it is capable of capturing some of the signals not detected by a single SNP association test(18).

In this study, we sought to identify biological pathways associated with MDD by making use of well-annotated molecular pathway databases and two independent samples of European ancestry. Each sample was run through a pipeline in which a non-hypothesis-driven pathway analysis was applied to identify MDD-associated pathways and that was followed by multi-level regional heritability analyses to quantify and narrow down the genetic contribution from the candidate pathways. Using this pipeline, we observed overlaps in the identified MDD-candidate pathways, genes and sub-regions between samples. Finally, in order to test for the predictive

value of the pathway most consistently associated with MDD across samples, we compared the predictive accuracy of pathway-derived MDD PRS to whole-genome-derived PRS.

## Methods and Materials

The Tayside Research Ethics Committee(reference 05/S1401/89) provided ethical approval for the study.

## Datasets

### *Generation Scotland: The Scottish Family Health Study (GS:SFHS)*

GS:SFHS included 21,387 subjects( $N_{\text{male}}=8,772$ ,  $N_{\text{female}}=12,615$ ;  $\text{Age}_{\text{mean}}=47.2$ ).

Participants were recruited from the registers of collaborating general practices via their Community Health Index(19). A structured clinical interview was used for the diagnosis of MDD DSM-IV mood disorders (SCID)(20)(see Supplement). By the time we performed this study, 9,863 individuals were genotyped using the Illumina Human OmniExpressExome -8- v1.0 array(21). Details of genotyping are described in detail elsewhere(22). Quality control(QC) and imputation method are described in the Supplement. In total, 592,690 genotyped and 2,163,848 imputed autosomal SNPs passed QC criteria and were used in subsequent analyses. Since close relatives can bias the pathway analysis and SNP heritability estimation, the function ‘--grm-cutoff 0.025’



in GCTA was used to remove one of each pair of individuals with estimated relatedness larger than 0.025 while maximizing the remaining sample size(11), 6,455 subjects (1,123 MDD cases and 5,332 controls) remained in the analyses described below.

*PGC Major Depression Dataset (PGC:MDD)*

The Psychiatric Genomics Consortium provided summary statistics from the GWAS mega-analysis of MDD from the discovery phase and individual genotypes from the nine primary cohorts in this dataset. These data included 18,759 subjects of European ancestry ( $N_{\text{MDD\_case}}=9,240$ ,  $N_{\text{control}}=9,519$ )(4). Cases were required to have a diagnosis of DSM-IV lifetime MDD (see Supplement). Summary statistics included GWAS P values and odds ratio information for 1,235,110 SNPs post-imputation(HapMap3). We carried out additional QC of these summary statistics using the following inclusion thresholds: info score  $\geq 0.8$ , and MAF  $\geq 0.01$ , after which 1,074,100 SNPs remained and were used in pathway analysis and polygenic risk profiling.

Imputed genotype data from nine PGC:MDD cohorts were provided by PGC for the regional heritability analysis. Best-guess imputed genotypes from each cohort were accepted at the same level of QC as GS:SFHS. After removing one of each pair of

close relatives( $t \geq 0.025$ ), the remaining 17,845 subjects were used in the downstream analyses(see Table s1 for sample information).

### **A pipeline for identification of pathways associated with MDD.**

This pipeline includes two stages of analyses: (1) a pathway analysis to identify MDD-associated pathways, and (2) multi-level (pathway/gene/sub-region) regional heritability analyses to narrow down the signals of the association. The multi-level regional heritability analyses for pathways identified by stage 1 were tested on the same sample in which they were first identified, and the test statistics from stage 2 are therefore potentially biased towards finding a more significant association. However, by applying the pipeline to two or more samples and by seeking replication across these the pipeline could provide independent replication of findings (as shown in current study). Further details of the pipeline are shown in Figure 1 and Table s2.

#### *Stage I: Pathway analysis*

For both samples, SNPs were annotated to 1,035 pathways(640 from Reactome, 216 from Biocarta and 179 from KEGG)(Supplement). For the GS:SFHS genotype dataset ( $N_{\text{SNP}}=592,690$ ;  $N_{\text{sample}}=6,455$ ), the “gene set ridge regression in association studies” algorithm (GRASS)(23) was used to identify pathway-MDD associations using only the genotyped SNPs(Supplement). False Discovery Rate (FDR)-adjusted P values

( $N_{FDR}=1035$ ) were calculated using the function `p.adjust` in the R package ‘stats’(24,25). For the PGC:MDD GWAS summary results dataset( $N_{SNP}=1,074,100$ ), Meta-Analysis Gene-set Enrichment of variant Associations(MAGENTA)(26) was used to test for the enrichment of genetic associations in each pathway for MDD, as just summary statistics were available for this part of the study (<http://www.med.unc.edu/pgc/downloads>) and MAGENTA was designed to exploit summary data from GWAS results(Supplement). MAGENTA reports a nominal P value and an estimated FDR per pathway( $N_{FDR}=1035$ ).

*Stage2: Estimation of MDD phenotypic variance explained by imputed genotypes of regional SNPs (regional heritability)*

Imputed SNPs were used in this analysis to avoid underestimating the regional heritability(12). For PGC:MDD, due to the heterogeneity caused by factors such as different ancestry and clinical diagnosis across samples(4)(Table s1), as well as analyzing the regional heritability in the combined dataset, we performed the SNP-heritability analysis in the three subsets used by the PGC:MDD consortium to group the nine cohorts (27)(Table s1).

We applied GREML using linear mixed modeling(LMM)(11) to estimate the variance explained by SNPs from genic regions of genes from candidate pathways and the sub-regions of candidate genes. A Log likelihood ratio test (LRT) was applied to test the significance of the estimated variance component(Supplement). Permutation analysis was performed to test whether the pathway-level regional heritability in candidate pathways was significantly greater than that expected by chance(Supplement). These analyses were performed in GCTA(11). To map regional heritability at sub-region level, regional heritability mapping(RHM, a modified GREML analysis) was applied using a sliding window to scan across the genic region of candidate genes(Supplement).

### **Polygenic profiling**

Polygenic risk scores (PRS)(28) estimate the genetic risk of MDD for unrelated individuals( $N_{\text{sample}}=6,455$ ) in GS:SFHS by adding the number of risk alleles an individual had, weighted by the effect size estimated in PGC:MDD(4)(Supplement).

To compare PRS derived from the pathway SNPs to that derived from the whole-genome SNPs, logistic regression and linear mixed models (LMM) were used to estimate the phenotypic variance explained by PRSs. (1)Logistic regression: PRS was treated as a fixed effect and MDD phenotype was regressed on PRS (other

covariates: age, age<sup>2</sup>, sex, top 4 principal components). The variance explained by PRS was calculated as Nagelkerke's  $R^2$  on the observed scale(29). Permutation analysis was conducted to set an empirical threshold by creating PRS from 1000 circularly permuted SNP sets(30)which were then fitted in logistic regression.

(2)LMM: we developed a PRS-bin-relationship-matrix method where the MDD phenotype variation was explained by the PRS similarity between subjects in the framework of LMM. The PRS similarity reflects relatedness in terms of the MDD-genetic risk and the variance explained by the PRS-similarity as a random effect was estimated using REML and tested using LRT(Supplement).

Finally, the area under the receiver-operating characteristic curve(AUC) was calculated to determine the efficacy of PRSs in correctly classifying MDD cases and controls using equation 2 in (31).

## Results

To identify candidate pathways for MDD, a pipeline that combines pathway and multi-level(pathway/gene/sub-region) regional heritability analyses was applied to two independent samples, GS:SFHS and PGC:MDD, respectively. Detailed information of the analytical pipeline and the data usage is shown in Figure 1 and Table s2.

### Identification of MDD-associated pathways in GS:SFHS

*Pathway analysis:* applying GRASS in GS:SFHS, four pathways were significantly associated with MDD after FDR correction ( $N_{FDR}=1035$ ). These comprised the following: three pathways from REACTOME (MTORC1 mediated signaling, NETRIN1 signaling, ABCA transporters in lipid homeostasis) and one pathway from BIOCARTEA (FEEDER pathway) (Table 1).

*Regional heritability analysis:* using GREML, the estimate of  $h_g^2$  (the heritability explained by all GWAS SNPs) for MDD was 0.25 (se=0.10) in GS:SFHS (Table s4A). To further investigate the regional heritability captured by SNPs from pathways that were significant in pathway analysis, for each pathway, we partitioned the genome-wide SNPs into two sets: (1) SNPs from the pathway and (2) the remaining SNPs. We then jointly estimated their contribution to MDD phenotypic variance in LMM. Among the four pathways that were significant in pathway analysis, two yielded significant P values (post-FDR-adjustment,  $N_{FDR}=4$ ) based on the LRT for pathway-level regional heritability in MDD, with the highest regional heritability estimated in the NETRIN1 signaling pathway ( $h_R^2 = 0.014$ , se=0.009,  $P_{lrt\_FDR}=0.019$ ) (Table 2). Permutation test across the circularly permuted SNP-sets with the same set size for the two pathways showed that the detected

pathway-regional-heritability was not attributable to gene set size and LD structures (NETRIN1 signaling:  $P_{perm} = 0.018$ ; MTORC1 mediated signaling pathway:  $P_{perm} = 0.01$ )(Supplement).

To narrow down the signals from the two pathways where significant pathway-level regional heritability was detected, gene-level regional heritability was estimated for single genes in the two pathways. The heritabilities for three genes (*DCC*, *UNC5D* and *SIAH2*) from the NETRIN1 pathway and one gene (*RPTOR*) from MTORC1 mediated signaling pathway obtained nominal significance in LRT (Table s5A). Among them, the receptor proteins encoded by *DCC* and *UNC5D* share the same ligand, Netrin-1, a key molecular signal in the NETRIN signaling pathway(32). To fine map the regional heritability within the two receptor genes and further explore if any of the sub-regions that conferred heritability overlapped with any functional domains, RHM was applied to the 2 genes using a fixed sliding window to scan across their genic regions (Figure 2, s1). Block 6 in *DCC* and Block 1 in *UNC5D* yield significance in LRT (Block6<sub>DCC</sub>:  $P_{lrt\_bonf} = 0.021$ ; Block1<sub>UNC5D</sub>:  $P_{lrt\_bonf} = 0.028$ )(Table s6A). Block 6 in *DCC* overlapped with the fourth Immunoglobulin-like domain (Ig)(Figure 2)(33). Block 1 in *UNC5D* overlapped with H3K4me3 signal region (Figure s1)(34).

### Identification of MDD-associated pathways in PGC:MDD

*Pathway analysis:* using MAGENTA, only 1 pathway from REACTOME (role of second messengers in NETRIN1 signaling. That is a subset of the ‘the NETRIN1 signaling pathway’ (100% overlap)(35)) was identified as associated with MDD after FDR correction( $N_{FDR}=1035$ ) (Table 1).

*Regional heritability analysis:* following the PGC published study, data from 9 cohorts were grouped into 3 subsets for the GREML analysis((27) and Table s1). The estimate of  $h_g^2$  for MDD varied from 0.26 (se=0.06) to 0.47 (se=0.05) across subsets (Table s4B). The pathway-level regional heritability from the role of second messengers in the NETRIN1 signaling pathway was nominally significant using LRT in subset1( $P_{lrt}=0.017$ ,  $P_{lrt\_FDR}=0.07$ ), while it was not significant in subset2, subset3 and the combined set(Table 2). Permutation test across the circularly permuted SNP-sets with the same set size in subset1 confirmed the enrichment of SNP-heritability in this pathway( $P_{perm}=0.01$ )(Supplement). The gene-level regional heritability analysis for this pathway obtained nominal significance in one gene (*DCC*,  $P_{lrt}=0.02$ ) in subset1, no genes in subset2, one gene (*TRPC3*) in subset3 and two genes(*PLCG1* and *PITPNA*) in the combined dataset (Table s5B). Since in subset1 the pathway-level regional heritability was significant for the role of second messengers in the NETRIN1 signaling pathway and the gene-level regional heritability was



significant in the *DCC* gene, we further localized the regional heritability by applying RHM to *DCC* in subset1 (Table s6B). The distribution of the regional heritability for the *DCC* gene in subset1 was similar to that obtained for in GS:SFHS (Figure 2).

### **Replication of GS:SFHS results in PGC:MDD and PGC:MDD results in GS:SFHS**

Among the four pathways identified by pathway analysis in GS:SFHS, the NETRIN1 signaling pathway was replicated in PGC:MDD ( $P_{path}=0.010$ ,  $N_{bonf}=4+1=5$ ,  $P_{path_{bonf}}=0.05$ ) whereas the other three pathways failed to replicate (Table 3). For the regional heritability analysis, among the 2 significant pathways in GS:SFHS, the NETRIN1 signaling pathway was significant in LRT in PGC:MDD subset1 ( $P_{lrt}=0.00258$ ,  $N_{bonf}=2*(1+3)+1=9$ ,  $P_{lrt_{bonf}}=0.02$ ), but not significant in other subsets or in the combined set (Table 4). The MTORC1 mediated signaling pathway failed to replicate in all subsets and in the combined set (Table 4). The gene-level regional heritability of *DCC* was nominally significant in both GS:SFHS (Table s5A) and in subset1 in PGC:MDD (Table s5B). The significant block 6 of *DCC* that was detected in GS:SFHS was fully covered by the nominally significant blocks in subset1 (Figure 2).

The only pathway identified by pathway analysis in PGC:MDD, the role of second messengers in NETRIN1 signaling pathway, was nominally significant in pathway analysis in GS:SFHS( $P_{path}=0.018$ ,  $N_{bonf}=5$ ,  $P_{path_{bonf}}=0.09$ )(Table 3). For the regional heritability analysis, this pathway was nominally significant in LRT in GS:SFHS( $P_{lrt}=0.017$ ,  $N_{bonf}=9$ ,  $P_{lrt_{bonf}}=0.156$ )(Table 4).

### **Estimation the predictive accuracy of Polygenic Risk Scores (PRS) derived from SNPs in the NETRIN1 signaling pathway or the whole genome**

We applied polygenic risk profiling to measure the additive genetic effect from the NETRIN1 signaling pathway and compared it with that from the whole genome(see methods).

Using logistic regression, for both pathway and whole-genome SNP sets, PRSs created without LD clumping explained a higher proportion of variance compared with PRSs created with LD clumping((Table s7, Figure 3). The PRS created from the whole genome SNPs explained a maximum MDD variance of 0.198%(GWAS  $P_{cutoff}=0.2$ , without LD clumping.  $P_{t-test}=0.006$ .). The PRS created from SNPs in the NETRIN1 signaling pathway explained a maximum variance of 0.216%(GWAS  $P_{cutoff}=0.2$ , without LD clumping.  $P_{t-test}=0.004$ .)(Table s7). Permutation test across

the circularly permuted SNP-sets with the same set size suggested that the variance explained by the NETRIN1 signaling pathway PRS (without LD clumping) in the logistic regression model was significantly higher than expected by chance (Table s8).

Using LMM, we estimated the proportion of MDD variance explained by pair-wise MDD-PRS similarity between individuals. A ‘PRS-bin relationship’ variance-covariance matrix was constructed and jointly fitted with a SNP-based genomic relationship matrix (GRM) in LMM. For each PRS, multiple bin numbers were tested to assess the stability of the model across different bin settings. When the comparison was between the PRSs created without LD clumping, the PRS-bin relationship matrices created from the NETRIN1 signaling pathway outperformed those created from the whole-genome set, as they consistently explained a significant proportion of variance across multiple bin settings and P value thresholds (Figure 4), with a maximum MDD variance explained of 1.7% ( $se=0.02$ ,  $P_{lrt}=0.013$ ,  $P_{perm}=0.012$ . PRS setting: bin=50, GWAS  $P_{cutoff}=0.5$ .) (Table s9). Nonetheless for PRSs created with LD clumping, most of the PRS-bin relationship matrices, both created from the whole genome SNPs or from the pathway SNPs, failed to obtain significance in LRT, with 3 exceptions at low-bin-setting (bin=10, 20, 50) for the whole-genome PRS-bin relationship matrix (Figure 4), the maximum MDD variance explained is 2.47% ( $se=0.028$ ,  $P_{lrt}=0.028$ ,  $P_{perm}=0.021$ . PRS setting: bin=10, GWAS  $P_{cutoff}=0.2$ ) (Table

s9). When jointly fitting PRS-bin relationship matrices from the pathway and whole genomes as well as a GRM in LMM, the variance explained by the NETRIN1 pathway PRS(without LD clumping) remained stable and significant(Table s10).

Finally, the estimation of the area under the ROC curve suggested that in general the AUCs of MDD PRSs were low, ranging from 0.498 to 0.532, with the NETRIN1 PRS obtaining the highest AUC( $AUC_{NETRIN1\_max}=0.532$ ,  $AUC_{whole\_max}=0.527$ )(Table s11).

## Discussion

Typically, disease-associated pathways have been assumed to have the following features: (1) the genetic variants in them are shown to be associated with disease in association tests; (2) the genetic variants in them explain a significant proportion of phenotypic variance; and (3) the genetic proxies such as polygenic risk scores derived from them have valuable predictive power for the disease. Here we applied a pipeline that integrates pathway analysis and regional heritability analyses to two independent samples. This enables the identification of candidate pathways for MDD that address the first two features. By comparing results from each stage of the pipeline, we identified the NETRIN1 signaling pathway which has multi-level associations with MDD that are observed across samples. Finally, the polygenic profiling method provided additional evidence that this pathway also satisfied the third feature.

In the pathway analysis, we identified four MDD-associated pathways in GS:SFHS but only one in PGC:MDD by using two methods, GRASS and MAGENTA, respectively, as different data types were available from each dataset for this analysis(Supplement). There was no pathway that was associated with MDD in both GS:SFHS and PGC:MDD. However the only associated pathway detected by MAGENTA in PGC:MDD, the role of second messengers in the NETRIN1 signaling pathway, was a subset of the NETRIN1 signaling pathway(100% overlap) which was detected by GRASS in GS:SFHS. Previous studies have suggested that the NETRIN1 signaling pathway plays a crucial role in axon guidance, a process that establishes precise brain circuits during the development of the central nervous system(36). Interestingly, in the development stage of the thalamus, the response of embryonic thalamocortical axons to the NETRIN1 signaling is modulated by serotonin(5-HT) signaling, a system that has been repeatedly implicated in the etiology of MDD(37,38). Given these convergent lines of research, NETRIN1 signaling is a promising candidate pathway for MDD.

The pathway analysis was followed by a pathway-level regional heritability analysis. We found that in both samples some of the MDD-associated pathways including the two NETRIN1 signaling pathways contributed significantly to explain MDD variance.

Moreover, the pathway-level regional heritability estimated was greater than that was expected given the SNP-set size in these pathways (Table 2), suggesting an enrichment of heritability further supported by the permutation test. These results were consistent with previous studies reporting enrichment of heritability in functionally annotated regions (12,13).

In the NETRIN1 signaling pathway, key proteins affect axon guidance: DCC is the key receptor for the attractive response to Netrin-1, whereas UNC5, alone or together with DCC, is associated with the repulsive response to Netrin-1 (36,39). In our study, the gene-based regional heritability analyses suggested that *DCC* and *UNC5D* were among the most associated genes in GS:SFHS, which is consistent with their functional importance in the NETRIN1 signaling pathway. In PGC:MDD, *DCC* was the only gene that attained nominal significance (in subset1, the only subset for which the pathway-level regional heritability from the candidate pathway is significant). When applying RHM to *DCC*, the regional heritability was localized to block6 in GS:SFHS (for PGC:MDD subset1 this region also obtains nominal significance) (Figure 2). Block6 overlapped with the fourth Immunoglobulin-like domain (Figure 2) which may be necessary for the axonal attraction mediated by Netrin-1 and draxin (40). A recent meta-analysis of GWAS for depressive symptom (N=180,886. The sample includes subjects from PGC MDD but the major source of

sample is from UK biobank( $N_{\text{ukb}}=105,739$ ) reported that one loci((rs62100776)) from the same gene *DCC* exceeded genome-wide significance(41). This SNP is located in block7, which is adjacent to the significant block6 and is nominal significant in our study(Figure 2). This overlapped finding using a much larger sample size validated our results and indirectly supported the value of applying our pipeline in studies with small sample size. The significant block in *UNC5D* overlapped with a H3K4me3 modification, which implicates an active promoter in that region(42). These results imply a potentially functional contribution of variants in *DCC* and *UNC5D* to MDD.

The pathway identified using our pipeline accounts for a significant proportion of phenotypic variance, which is an attractive feature for a biomarker. Polygenic risk scores(PRSs) were created by adding-up the genetic effects among biomarkers. Although MDD is a highly polygenic disorder, PRSs from whole genome SNPs can be noisy as they include SNPs with no effect on MDD. A more accurate prediction is likely from scores derived from biomarkers with a higher proportion of causal SNPs. We thus measured the prediction value for MDD of the PRSs derived from the associated NETRIN1 signaling pathway. The results showed that pathway-PRSs explain a higher proportion of phenotypic variance than whole-genome-PRSs, when PRS was fitted as a fixed effect in the logistic regression. The AUC statistics also

support a better prediction by the pathway-PRSs. We additionally developed a ‘PRS-bin-relationship matrix’ method in which PRS-similarity was used to explain phenotypic variation of MDD in LMM. Using this method, the MDD phenotypic variance explained by the pathway-PRS and whole-genome-PRS was substantially increased to 1.70% and 2.47%, respectively. Although the largest variance explained by the pathway-PRS was smaller than the whole-genome-PRS, the pathway-PRS performed better in terms of the significance level in LRT across most of the tested bins(Figure 4).

Notably, although the ‘PRS-bin relationship matrix’ is conceptually similar to the classic common-SNP-based GRM, a key difference was that the ‘PRS-bin relationship matrix’ took the information of the effect size of loci as estimated in the discovery sample, PGC:MDD, and the genotypes from the target sample, GS:SFHS, so that it measures the MDD-genetic-risk-similarity and the interpretation of the model which it fitted was across samples. This method also enabled the discrimination of genetic effects represented in the pathway-PRS, whole-genome-PRS and GRM when they were jointly fitted in LMM. Our results suggest that they explained distinct proportions of phenotypic variance(1.6% for pathway-PRS, 2.2% for whole-genome-PRS and 22.7% for whole-genome-GRM)(Table s10).



Limitations and further research: firstly, different methods were applied in different samples using different data formats in the pathway analysis. This may have influenced the consistency and the comparability of the results, and complicated their interpretation. Secondly, it is possible that the predictive accuracy of the NETRIN1 signaling pathway measured in this study is inflated, as both samples were involved in the pipeline where this pathway was identified(albeit independently), further replication in independent populations would strengthen our findings. Note that although the multi-level regional-heritability analyses enable the fine-mapping of association signals, their P-values may be inflated as the fine-mapping was conducted at the same datasets where the pathways were identified. We therefore suggested a carefully interpretation of the LRT results from the regional heritability analyses, and we attach more importance to the comparisons of the regional heritability patterns across populations(Figure 2). Thirdly, the route by which NETRIN1 signaling pathway contributes to MDD is unknown. Future directions for increasing our understanding could include (1)exploring mutant DCC animal models of MDD, (2)testing of the interactions of NETRIN1-receptors with known MDD-associated proteins.

In summary, this study shows that the NETRIN1 signaling pathway was associated with MDD in two independent samples. Variants in this pathway accounted for a

significant proportion of variance in susceptibility for MDD, and have valuable prediction power. These findings further support a role for NETRIN1 in the etiology of MDD and provide a basis for future studies.

Accepted manuscript

## Acknowledgements

We are grateful to the families who took part in GS:SFHS, the GPs and Scottish School of Primary Care for their help in recruiting them, and the whole GS team, which includes academic researchers, clinic staff, laboratory technicians, clerical workers, IT staff, statisticians and research managers. YNZ is grateful for Mr. Ian White's suggestion for analysis of polygenic score. This work is supported by the Wellcome Trust through a Strategic Award, reference 104036/Z/14/Z. The Chief Scientist Office of the Scottish Government and the Scottish Funding Council provided core support for Generation Scotland. GS:SFHS was funded by a grant from the Scottish Government Health Department, Chief Scientist Office, number CZD/16/6. The authors TKC and AMMcI acknowledge with gratitude the financial support received for this work from the Dr Mortimer and Theresa Sackler Foundation. PAT, DJP, IJD and AMMcI are members of The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. DJM is an NRS Fellow, funded by the CSO. PN and CSH acknowledge support from the MRC to the HGU and from the BBSRC (BB/J004235/1).

Membership List of Major Depressive Disorder Working Group of the Psychiatric

GWAS Consortium (PGC):

Stephan Ripke, Naomi R. Wray, Cathryn M. Lewis, Steven P. Hamilton, Myrna M. Weissman, Gerome Breen, Enda M. Byrne, Douglas H.R. Blackwood, Dorret I. Boomsma, Sven Cichon, Andrew C. Heath, Florian Holsboer, Susanne Lucae, Pamela A.F. Madden, Nicholas G. Martin, Peter McGuffin, Pierandrea Muglia, Markus M. Nothen, Brenda P. Penninx, Michele L. Pergadia, James B. Potash, Marcella Rietschel, Danyu Lin, Bertram Müller-Myhsok, Jianxin Shi, Stacy Steinberg, Hans J. Grabe, Paul Lichtenstein, Patrik Magnusson, Roy H. Perlis, Martin Preisig, Jordan W. Smoller, Kari Stefansson, Rudolf Uher, Zoltan Kutalik, Katherine E. Tansey, Alexander Teumer, Alexander Viktorin, Michael R. Barnes, Thomas Bettecken, Elisabeth B. Binder, René Breuer, Victor M. Castro, Susanne E. Churchill, William H. Coryell, Nick Craddock, Ian W. Craig, Darina Czamara, Eco J. De Geus, Franziska Degenhardt, Anne E. Farmer, Maurizio Fava, Josef Frank, Vivian S. Gainer, Patience J. Gallagher, Scott D. Gordon, Sergey Goryachev, Magdalena Gross, Michel Guipponi, Anjali K. Henders, Stefan Herms, Ian B. Hickie, Susanne Hoefels, Witte Hoogendijk, Jouke Jan Hottenga, Dan V. Iosifescu, Marcus Ising, Ian Jones, Lisa Jones, Tzeng Jung-Ying, James A. Knowles, Isaac S. Kohane, Martin A. Kohli, Ania Korszun, Mikael Landen, William B. Lawson, Margarita Rivera, Glyn Lewis, Donald MacIntyre, Wolfgang Maier, Manuel Mattheisen, Patrick J. McGrath, Andrew M. McIntosh, Alan McLean, Christel M. Middeldorp, Lefkos Middleton, Kloiber S, Grant M. Montgomery, Shawn N. Murphy, Matthias Nauck, Willem A. Nolen, Dale R.

Nyholt, Michael O'Donovan, Högni Oskarsson, Nancy Pedersen, William A. Scheftner, Andrea Schulz, Thomas G. Schulze, Stanley I. Shyn, Engilbert Sigurdsson, Susan L. Slager, Johannes H. Smit, Hreinn Stefansson, Michael Steffens, Thorgeir Thorgeirsson, Federica Tozzi, Jens Treutlein, Manfred Uhr, Edwin J.C.G. van den Oord, Gerard Van Grootheest, Henry Völzke, Jeffrey B. Weilburg, Gonneke Willemsen, Frans G. Zitman, Benjamin Neale, Mark Daly, Douglas F. Levinson, Patrick F. Sullivan

### **Financial Disclosures**

AMM has previously received grant support from Pfizer, Lilly and Janssen. These studies are not connected to the current investigation. This work is supported by the Wellcome Trust through a Strategic Award, reference 104036/Z/14/Z. The Chief Scientist Office of the Scottish Government and the Scottish Funding Council provided core support for Generation Scotland. GS:SFHS was funded by a grant from the Scottish Government Health Department, Chief Scientist Office, number CZD/16/6. The author YNZ acknowledge the support from China Scholarship Council. The authors TKC and AMMcI acknowledge with gratitude the financial support received for this work from the Dr Mortimer and Theresa Sackler Foundation. PAT, DJP, IJD and AMMcI are members of The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong

Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. DJM is an NRS Fellow, funded by the CSO. PN and CSH acknowledge support from the MRC. All other authors report no biomedical financial interests or potential conflicts of interest.

Accepted manuscript

## Legends

**Table 1** Top 10 pathways in pathway analysis for MDD using GRASS on GS:SFHS and using MAGENTA on PGC:MDD.  $N_{FDR} = 1035$ . For GRASS results on GS:SFHS, four pathways yielded significance after FDR correction. For MAGENTA results on PGC:MDD, one pathway yielded significance after FDR correction. EXP # GENES ABOVE 95% CUTOFF: expected number of genes with a corrected gene P value above the 95 percentile enrichment cutoff. OBS # GENES ABOVE 95% CUTOFF: observed number of genes with a corrected gene P value above the 95 percentile enrichment cutoff.

**Table 2** Pathway-level regional heritability analysis results for the significant pathways identified in pathway analysis for GS:SFHS and PGC:MDD.  $h_R^2$ : pathway-level regional heritability attributable to the pathway SNPs.  $h_C^2$ : heritability attributable to the complement SNP set.  $LRT(h_R^2)$  P value and  $LRT(h_R^2)$   $P_{FDR}$ : Nominal P value and FDR adjusted P value from LRT for  $h_R^2$ .  $N_{pathway\_snp}$ : SNP number in the pathway.  $\%h_{gwas}^2/\%snp$ : the ratio of the percentage of  $h_{gwas}^2$  in pathway to the percentage of SNPs in the pathway. \* - significant results after multiple test correction (5%).

**Table 3.** Replication results of pathway analysis. Four pathways were identified from GS:SFHS in PGC:MDD and one pathway was identified from PGC:MDD in

GS:SFHS.  $P_{\text{bonf}}$ : adjusted P value using Bonferroni multiple testing correction.  $N_{\text{bonf}}=5$ .

**Table 4.** Replication results of pathway-level regional heritability analysis. Two pathways were significant in GS:SFHS in PGC:MDD, and one pathway was significant in PGC:MDD in GS:SFHS. **LRT( $h^2_R$ ) P value:** adjusted P value using Bonferroni multiple testing correction.  $N_{\text{bonf}}=9$ .

**Figure 1.** The analytical pipeline, its application in identifying associated pathways with MDD and its findings in two independent samples: GS:SFHS and PGC:MDD. A: Design of the analytical pipeline. In the pipeline, pathway analysis was performed for 1035 pathways, significant pathways were analysed using multi-level regional heritability analyses(RHA) in the framework of Genomic Restricted Maximum Likelihood analysis(GREML) to quantify and localize the genetic effects on MDD. In pathway analysis, GRASS was applied to the phenotype and genotype data of 6,455 individuals from the GS:SFHS sample. MAGENTA was applied to the summary data from PGC MDD GWAS on 18,659 individuals. B: The findings by the analytical pipeline in the two samples. The NETRIN1 signaling pathway was identified as associated pathway with MDD and the association signal were localized to its gene *DCC* and the sub-region level.



**Figure 2** Genic region in *DCC* showing blocks used in REACTA on GS:SFHS and subset1 in PGC:MDD subset1(subsets 2 and 3 failed to obtained significance in the pathway-level regional heritability of NETRIN1 signaling pathway). In GS:SFHS, the sliding window (window size=200 SNPs) defined 9 blocks with average block size of 179kb. In PGC:MDD subset1, window size=100 SNPs was used as the density of SNPs in PGC:MDD dataset was around half of that in GS:SFHS. This divided *DCC* into 8 blocks with average block size of 191kbs. Blue bar: insignificant region in LRT. Orange bar: significant region in LRT. Red bar: significant region in LRT after Bonferroni correction. Red dotted line: significant Block 6 in GS:SFHS, which overlaps with the fourth Ig (Immunoglobulin-like) domain. This region was fully covered by the nominal significant regions in subset1 in PGC:MDD.

**Figure 3.** The phenotypic variance explained by polygenic risk score as a fixed effect in logistic regression. Perm Aver NETRIN1 NETRIN1 without LD clumping: the average Nagelkerke's  $R^2$  of 1000 PRSs created from permuted pathway SNPs (the circular permuted SNP-sets with the same set size).

**Figure 4.** The LRT result from LMM showing the significance level of the phenotypic variance explained by PRS-bin relationship matrices derived from variants in whole genome and the NETRIN1 signaling pathway with or without LD clumping

at GWAS P value thresholds of 0.2, 0.5, and 1 using different bins. The color of the bars was designated by the bin number. Red line marks the significance level ( $P_{lri}=0.05$ ).

Accepted manuscript

## References

1. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. (2013) Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med* 10: e1001547.
2. Sullivan PF, Neale MC, Kendler KS (2000) Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 157: 1552-1562.
3. consortium C (2015) Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 523: 588-591.
4. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 18: 497-511.
5. Schizophrenia Working Group of the Psychiatric Genomics C (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511: 421-427.
6. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, et al. (2013) All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet* 9: e1003449.
7. Sullivan PF, Posthuma D (2015) Biological pathways and networks implicated in psychiatric disorders. *Current Opinion in Behavioral Sciences* 2: 58-68.
8. Network, Pathway Analysis Subgroup of Psychiatric Genomics C (2015) Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 18: 199-209.
9. Ramanan VK, Shen L, Moore JH, Saykin AJ (2012) Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet* 28: 323-332.
10. Juraeva D, Haenisch B, Zapatka M, Frank J, Investigators G, group P-GSw, et al. (2014) Integrated Pathway-Based Approach Identifies Association between Genomic Regions at CTCF and CACNB2 and Schizophrenia. *PLoS Genet* 10: e1004345.
11. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88: 76-82.
12. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson Bjarni J, Xu H, et al. (2014) Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *The American Journal of Human Genetics* 95: 535-552.
13. Torres Jason M, Gamazon Eric R, Parra Esteban J, Below Jennifer E, Valladares-Salgado A, Wacher N, et al. (2014) Cross-Tissue and Tissue-Specific eQTLs: Partitioning the Heritability of a Complex Trait.

- The American Journal of Human Genetics 95: 521-534.
14. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* 43: 519-U544.
  15. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM (2015) Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet* 11: e1004969.
  16. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*.
  17. Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I, et al. (2012) Localising loci underlying complex trait variation using Regional Genomic Relationship Mapping. *PLoS One* 7: e46501.
  18. Riggio V, Matika O, Pong-Wong R, Stear MJ, Bishop SC (2013) Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity (Edinb)* 110: 420-429.
  19. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. (2013) Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology* 42: 689-700.
  20. First MB, Spitzer RL, Gibbon M, Williams JB (2012) Structured Clinical Interview for DSM-IV® Axis I Disorders (SCID-I), Clinician Version, Administration Booklet: American Psychiatric Pub.
  21. Gunderson KL (2009) Whole-genome genotyping on bead arrays. *Methods Mol Biol* 529: 197-213.
  22. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. (2006) Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet* 7: 74.
  23. Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, et al. (2010) Insights into Colon Cancer Etiology via a Regularized Approach to Gene Set Analysis of GWAS Data. *American Journal of Human Genetics* 86: 860-871.
  24. Team RC (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.
  25. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 57: 289-300.
  26. Segre AV, Consortium D, investigators M, Groop L, Mootha VK, Daly MJ, et al. (2010) Common inherited variation in mitochondrial genes is not

- enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 6.
27. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* 45: 984-+.
  28. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460: 748-752.
  29. Lee SH, Goddard ME, Wray NR, Visscher PM (2012) A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* 36: 214-224.
  30. Cabrera CP, Navarro P, Huffman JE, Wright AF, Hayward C, Campbell H, et al. (2012) Uncovering networks from genome-wide association studies via circular genomic permutation. *G3 (Bethesda)* 2: 1067-1075.
  31. Wray NR, Yang J, Goddard ME, Visscher PM (2010) The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet* 6: e1000864.
  32. Mehlen P, Mazelin L (2003) The dependence receptors DCC and UNC5H as a link between neuronal guidance and survival. *Biology of the Cell* 95: 425-436.
  33. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38: D211-222.
  34. Consortium EP (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306: 636-640.
  35. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res* 42: D472-477.
  36. Moore SW, Tessier-Lavigne M, Kennedy TE (2007) Netrins and their receptors. *Axon Growth and Guidance* 621: 17-31.
  37. Bonnin A, Torii M, Wang L, Rakic P, Levitt P (2007) Serotonin modulates the response of embryonic thalamocortical axons to netrin-1. *Nat Neurosci* 10: 588-597.
  38. Jans LA, Riedel WJ, Markus CR, Blokland A (2007) Serotonergic vulnerability and depression: assumptions, experimental evidence and implications. *Mol Psychiatry* 12: 522-543.
  39. TessierLavigne M, Goodman CS (1996) The molecular biology of axon guidance. *Science* 274: 1123-1133.
  40. Chen Q, Sun XQ, Zhou XH, Liu JH, Wu JN, Zhang Y, et al. (2013) N-terminal horseshoe conformation of DCC is functionally required for axon guidance and might be shared by other neural receptors. *Journal of Cell Science* 126: 186-195.
  41. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, et al. (2016) Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses.

Nat Genet.

42. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, et al. (2008)  
Divergent Transcription from Active Promoters. Science 322: 1849-1851.

Accepted manuscript

**Table 1.**

| Ra<br>nk       | Datab<br>ase | Pathway   | P<br>value   | P <sub>FDR</sub> | Eff Gene<br>Size | Exp # Genes Above<br>95% Cutoff | Obs # Genes Above<br>95% Cutoff |
|----------------|--------------|---|--------------|------------------|------------------|---------------------------------|---------------------------------|
| <b>GS:SFHS</b> |              |   |              |                  |                  |                                 |                                 |
| 1              | BioCa<br>rta | Feeder<br>pathway   | 2.93E<br>-05 | 2.14E-<br>02 *   | -                | -                               | -                               |
| 2              | Reacto<br>me | ABCA<br>transport<br>ers in<br>lipid<br>homeost<br>asis                 | 5.20E<br>-05 | 2.14E-<br>02 *   | -                | -                               | -                               |
| 3              | Reacto<br>me | NETRIN<br>1<br>signaling  | 6.20E<br>-05 | 2.14E-<br>02 *   | -                | -                               | -                               |
| 4              | Reacto<br>me | MTORC<br>1<br>mediated<br>signallin<br>g                                | 8.87E<br>-05 | 2.30E-<br>02 *   | -                | -                               | -                               |
| 5              | BioCa<br>rta | P35<br>Alzheim<br>er's<br>pathway                                       | 2.69E<br>-04 | 5.58E-<br>02     | -                | -                               | -                               |
| 6              | BioCa<br>rta | SODD<br>pathway   | 5.23E<br>-04 | 8.59E-<br>02     | -                | -                               | -                               |
| 7              | Reacto<br>me | Energy<br>depende<br>nt<br>regulatio<br>n of<br>MTOR<br>by LKB1<br>AMPK | 7.46E<br>-04 | 8.59E-<br>02     | -                | -                               | -                               |
| 8              | Reacto<br>me | NFKB<br>activatio<br>n<br>through<br>FADD<br>RIP1                       | 7.47E<br>-04 | 8.59E-<br>02     | -                | -                               | -                               |

|                |          |  |          |           |    |   |   |
|----------------|----------|--|----------|-----------|----|---|---|
|                |          | pathway mediated by caspase 8 and 10                   |          |           |    |   |   |
| 9              | KEGG     | Taste transduction                                     | 6.48E-04 | 8.59E-02  | -  | - | - |
| 10             | Reactome | ABC family proteins mediated transport                 | 8.68E-04 | 8.98E-02  | -  | - | - |
| <b>PGC:MDD</b> |          |  |          |           |    |   |   |
| 1              | Reactome | Role of second messengers in NETRIN 1 signaling        | 1.00E-04 | 1.46E-02* | 8  | 0 | 4 |
| 2              | Reactome | Defensins  | 9.00E-04 | 2.04E-01  | 25 | 1 | 6 |
| 3              | Reactome | NRAGE signals death through JNK                        | 1.50E-03 | 2.07E-01  | 36 | 2 | 7 |
| 4              | Reactome | Beta defensins   | 3.50E-03 | 2.75E-01  | 21 | 1 | 5 |
| 5              | Reactome | Purine catabolism                                      | 1.26E-02 | 3.11E-01  | 10 | 1 | 3 |
| 6              | Reactome | Formation of tubulin folding intermediates by CCT TRIC | 1.71E-02 | 6.67E-01  | 22 | 1 | 4 |



|    |          |   |          |          |    |   |   |
|----|----------|---|----------|----------|----|---|---|
| 7  | BioCarta | AKAP13 pathway                                  | 1.88E-02 | 7.31E-01 | 12 | 1 | 3 |
| 8  | Reactome | Chondroitin sulfate dermatan sulfate metabolism | 5.54E-02 | 7.41E-01 | 49 | 2 | 5 |
| 9  | Reactome | Elevation of cytosolic CA2 levels               | 5.67E-02 | 7.43E-01 | 10 | 0 | 2 |
| 10 | Reactome | Opsins  | 4.30E-02 | 7.46E-01 | 10 | 0 | 2 |

Table 2

| Pathway or Group   | $h^2_R$ | $se(h^2_R)$ | $h^2_C$ | $se(h^2_C)$ | LRT( $h^2_R$ )<br>P value | LRT( $h^2_R$ )<br>P <sub>FDR</sub> | N <sub>pathway</sub><br>SNPs | %snp  | % $h^2_{gwas}$ | % $h^2_{gwas}/\%$ snp |
|--|---------|-------------|---------|-------------|---------------------------|------------------------------------|------------------------------|-------|----------------|-----------------------|
| <b>GS:SFHS</b>   |         |             |         |             |                           |                                    |                              |       |                |                       |
| Reactome MTO RC1 Mediated Signaling                            | 0.006   | 0.004       | 0.240   | 0.099       | 7.70E-03                  | 1.86E-02*                          | 947                          | 0.04  | 2.40           | 54.84                 |
| Reactome NETRIN1 Signaling                                     | 0.014   | 0.009       | 0.224   | 0.099       | 9.28E-03                  | 1.86E-02*                          | 8809                         | 0.41  | 5.80           | 14.20                 |
| BioCarta Feeder Pathway  | 0.004   | 0.004       | 0.251   | 0.099       | 3.77E-02                  | 5.03E-02                           | 507                          | 0.02  | 1.00           | 42.68                 |
| Reactome ABCA Transporters In Lipid Homeostasis                | 0.000   | 0.004       | 0.251   | 0.099       | 5.00E-01                  | 5.00E-01                           | 1020                         | 0.05  | 0.00           | 0.00                  |
| <b>PGC:MDD: Role Of Second Messengers In NETRIN1 Signaling</b> |         |             |         |             |                           |                                    |                              |       |                |                       |
| PGC: MDD combined  | 0.0001  | 0.0006      | 0.2616  | 0.0216      | 3.77E-01                  | 4.17E-01                           | 1083                         | 0.001 | 0.0005         | 0.52                  |

|                            |            |            |            |            |              |               |      |           |            |       |
|----------------------------|------------|------------|------------|------------|--------------|---------------|------|-----------|------------|-------|
| PGC:<br>MDD<br>subset<br>1 | 0.00<br>55 | 0.00<br>34 | 0.26<br>25 | 0.05<br>96 | 9.06E<br>-03 | 3.62E-<br>02* | 1083 | 0.0<br>01 | 0.01<br>93 | 19.12 |
| PGC:<br>MDD<br>subset<br>2 | 0.00<br>10 | 0.00<br>18 | 0.46<br>55 | 0.05<br>25 | 2.56E<br>-01 | 4.17E-<br>01  | 1083 | 0.0<br>01 | 0.00<br>34 | 3.37  |
| PGC:<br>MDD<br>subset<br>3 | 0.00<br>04 | 0.00<br>22 | 0.44<br>42 | 0.07<br>80 | 4.17E<br>-01 | 4.17E-<br>01  | 1083 | 0.0<br>01 | 0.00<br>14 | 1.41  |

**Table 3**

| Datab<br>ase              | Pathwa<br>y   | P<br>value   | P <sub>bonf</sub> | Eff GENE<br>SIZE | EXP # GENES ABOVE<br>95% CUTOFF | OBS # GENES ABOVE<br>95% CUTOFF |
|---------------------------|---|--------------|-------------------|------------------|---------------------------------|---------------------------------|
| <b>GS:SFHS in PGC:MDD</b> |   |              |                   |                  |                                 |                                 |
| React<br>ome              | ABCA<br>transpor<br>ters in<br>lipid<br>homeos<br>tasis | 5.39E<br>-01 | 1.00E+<br>00      | 15               | 1                               | 1                               |
| React<br>ome              | NETRI<br>N1<br>signalin<br>g                            | 9.90E<br>-03 | 4.95E-<br>02 *    | 37               | 2                               | 6                               |
| React<br>ome              | MTOR<br>C1<br>mediate<br>d<br>signalli<br>ng            | 4.05E<br>-01 | 1.00E+<br>00      | 10               | 1                               | 1                               |
| BioCa                     | Feeder<br>pathwa  | 1.00E        | 1.00E+            | 9                | 0                               | 0                               |

|                           |   |          |          |   |   |   |
|---------------------------|---|----------|----------|---|---|---|
| rt                        | y   | +00      | 00       |   |   |   |
| <b>PGC:MDD in GS:SFHS</b> |   |          |          |   |   |   |
| Reactome                  | Role of second messengers in NETRI N1 signaling | 1.75E-02 | 8.76E-02 | - | - | - |

Accepted manuscript

**Table 4.**

| GS:SFHS in<br>PGC:MDD:<br>Group                         | NETRIN1 Signaling |                   |           |                   |                           |                                  | MTORC1 Mediated Signalling |               |            |               |                           |                                  |
|---|-------------------|-------------------|-----------|-------------------|---------------------------|----------------------------------|----------------------------|---------------|------------|---------------|---------------------------|----------------------------------|
|   | $h^2_R$           | Se<br>( $h^2_R$ ) | $h^2_C$   | Se<br>( $h^2_C$ ) | LRT( $h^2_R$ ) P<br>value | LRT( $h^2_R$ ) $P_{\text{bonf}}$ | $h^2_R$                    | se( $h^2_R$ ) | $h^2_C$    | se( $h^2_C$ ) | LRT( $h^2_R$ ) P<br>value | LRT( $h^2_R$ ) $P_{\text{bonf}}$ |
| Combined  | 0.00<br>0         | 0.00<br>1         | 0.26<br>2 | 0.02<br>2         | 4.75E-<br>01              | 1.00E+<br>00                     | 0.000<br>3                 | 0.00<br>05    | 0.28<br>43 | 0.02<br>20    | 2.71E-<br>01              | 1.00E+00                         |
| Subset1   | 0.01<br>4         | 0.00<br>6         | 0.25<br>6 | 0.05<br>9         | 2.59E-<br>03              | 2.33E-0<br>2*                    | 0.000<br>6                 | 0.00<br>15    | 0.26<br>47 | 0.05<br>97    | 3.32E-<br>01              | 1.00E+00                         |
| Subset2   | 0.00<br>3         | 0.00<br>4         | 0.46<br>3 | 0.05<br>2         | 1.59E-<br>01              | 1.00E+<br>00                     | 0.001<br>8                 | 0.00<br>19    | 0.46<br>43 | 0.05<br>24    | 1.02E-<br>01              | 9.17E-01                         |
| Subset3   | 0.00<br>2         | 0.00<br>5         | 0.44<br>3 | 0.07<br>8         | 3.83E-<br>01              | 1.00E+<br>00                     | 0.000<br>0                 | 0.00<br>18    | 0.44<br>47 | 0.07<br>81    | 5.00E-<br>01              | 1.00E+00                         |
| PGC:MDD in GS:SFHS: Pathway                             |                   |                   |           |                   |                           |                                  | $h^2_R$                    | se( $h^2_R$ ) | $h^2_C$    | se( $h^2_C$ ) | LRT( $h^2_R$ ) P<br>value | LRT( $h^2_R$ ) $P_{\text{bonf}}$ |
| Reactome Role Of Second Messengers In NETRIN1 Signaling |                   |                   |           |                   |                           |                                  | 0.005                      | 0.00<br>4     | 0.23<br>5  | 0.09<br>9     | 1.73E-<br>02              | 1.56E-01                         |

